

COVERING SYSTEM FOR LANGUAGE TREE PATTERN OF MACHINE TRANSLATION

Publication number: JP60247784

Publication date: 1985-12-07

Inventor: SAKAKI HIROSHI; HASHIMOTO KAZUO; SUZUKI MASAMI; YAGAKINAI IZURU

Applicant: KOKUSAI DENSHIN DENWA CO LTD

Classification:

- International: G06F17/27; G06F17/28; G06F17/27; G06F17/28; (IPC1-7): G06F15/38

- European: G06F17/27A

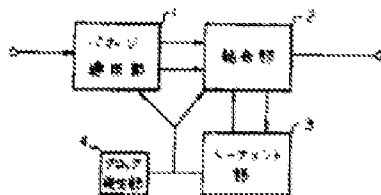
Application number: JP19840102651 19840523

Priority number(s): JP19840102651 19840523

Report a data error here

Abstract of JP60247784

PURPOSE:To cover a given tree in an optimum state by considering the opposition possibility of a pattern when the given tree is covered from above. **CONSTITUTION:**A pattern application part 1 covers the given tree with the pattern read out of its pattern dictionary memory from the uppermost part. Then, a tree having a partial tree which is not covered with the pattern and a tree determined by being covered with the pattern are transferred to a coupling part 2. The coupling part 2 performs arithmetic at its coupling arithmetic part and said tree having the uncovered partial tree is covered successively with the determined tree. Then, when an input signal has opposition possibility, i.e. includes two partial parts, it is supplied to a tournament part 3 calculates the low-order total cost function of respective partial trees to decide a winner, and the input tree is determined and outputted to the coupling part 2.



⑫ 公開特許公報(A)

昭60-247784

⑬ Int.Cl.⁴

識別記号

庁内整理番号

⑭ 公開 昭和60年(1985)12月7日

G 06 F 15/38

7313-5B

審査請求 未請求 発明の数 1 (全19頁)

⑮ 発明の名称 機械翻訳における言語木パターン被覆方式

⑯ 特 願 昭59-102651

⑰ 出 願 昭59(1984)5月23日

⑱ 発 明 者 榊 博 史 東京都目黒区中目黒2丁目1番23号 国際電信電話株式会社研究所内

⑲ 発 明 者 橋 本 和 夫 東京都目黒区中目黒2丁目1番23号 国際電信電話株式会社研究所内

⑳ 発 明 者 鈴木 雅 実 東京都目黒区中目黒2丁目1番23号 国際電信電話株式会社研究所内

㉑ 発 明 者 野 垣 内 出 東京都目黒区中目黒2丁目1番23号 国際電信電話株式会社研究所内

㉒ 出 願 人 国際電信電話株式会社 東京都新宿区西新宿2丁目3番2号

㉓ 代 理 人 弁理士 山本 恵一

明 細 書

1. 発明の名称

機械翻訳における言語木パターン被覆方式

2. 特許請求の範囲

(1) 電子手段により言語処理における木の一部又は全部を少なくともひとつのパターンで重複なくかつ間隙なく覆い、かつ、パターンの再上位節点で覆われる節点は重複して覆う方式において、与えられた木の最上位部分より下位方向に向って被覆可能なパターンで前記条件に従って木を覆い記憶する第1ステップと、被覆したパターンについて木の下位より上位方向に向ってパターン毎に定められるコスト関数に従ってコスト関数の総和が最大となるごとくトーナメントのパスを定める第2ステップとを有し、与えられたコスト体系のもとで最適な木の被覆方法を決定することを特徴とする言語木パターン被覆方式。

(2) 2以上の木が与えられたとき、各木毎の最適被覆方法によるコスト関数の値が最大の木を最適な木として選択することを特徴とする特許請求

の範囲第1項記載の言語木パターン被覆方式。

3. 発明の詳細な説明

(技術分野)

本発明は機械翻訳方式に係り、特に、与えられた言語木を、与えられた多種類の木状パターンで重複かつ間隙なく覆うための言語木の被覆方式に関する。

(背景技術)

従来の機械翻訳方式の一つに、言語木を多くの特定の木状パターンの集合ととらえ、各パターン毎にそれに対応する翻訳先の言語、すなわち目的言語のパターンを与えることにより翻訳を行うものがある(例えば「特開昭58-201175 機械翻訳方式」)。

先ず、この従来技術について、図1～図6を用い簡単に説明する。翻訳される文が、「I play the piano.」という英語文であり、これが日本語に翻訳される例である。

入力文は、大要、次の6つの手続きを経て翻訳される。

手続 1: 単語名を節点名とする節点に分離し、節点の集合を構成する(図1)。この節点は、後述する言語木(または単に木)の終端節点と呼ばれるものになる。

手続 2: 単語名を節点とする各節点の品詞名を確定し、品詞名を節点名とする節点を発生し、対応する節点を同志を結ぶ(図2)。ここで、節点と節点を結ぶ線は「枝」と呼ばれる。

手続 3: 「同一層上のいくつかの節点をまとめて、一つの上位節点に結ぶ。」というパーシング文法を適用し、これを繰返すことにより、被分離の言語木を構成する(図3)。木とは、ある節点から出発して、異なる枝と節点をたどったのでは元の節点に戻る経路の無いものを指す。

図3において、Tはテキスト、Sは文、NPは名刺句、NPOは人称代名詞、NMIは楽器名、VTは他動詞、DETは限定詞、ENDは終止記号を表わす。

ある段階で不適切な木状パターンが選択されることが発生し、下方に手続が進むと木を覆うことのできる木状パターンが無くなってしまったり、不適切な目的言語を与えていた、ということである。この点について、さらに述べる。なお、以降の説明では図3に相当する木を「与えられた木」と呼び(手続3から手続4へ与えられた、という意味で)、また、木状パターンを単にパターンと呼ぶ。また、与えられた木を図7に示すように、節点名を特に意味を持たない記号で表現することで一般化し、説明を進める。パターンも同様に表現し、図8～図19に示す12種類が準備されているものとし、図8をパターンE、図9をパターンF、図10をパターンI、図11をパターンL、図12をパターンM、図13をパターンN、図14をパターンR、図15をパターンS、図16をパターンT、図17をパターンX、図18をパターンY、図19をパターンZと呼ぶこととする。

図7の与えられた木を、その上方からパターンで覆う場合、最初の手続で、図20～図22に示

手続 4: 木に対して、上方から下方に向けて、与えられた多種類の木状パターンのうちから適当な木状パターンを選択し、複数の木状パターンで木全体を覆う(図5)。図5の例では、選択された木状パターンは図4(a)～図4(i)である。木を覆う規則は、「各木状パターンは唯一の最上位節点を持ち、この最上位節点が上位の木状パターン最下位節点の一つを構成する。」というものである。

手続 5: 木を覆うために用いた木状パターンの種類と組合せなどから、助詞J'や助動詞JD'を追加しつつ、各木状パターンに目的言語を与える(図6)。

手続 6: 全ての終端節点を連結し、翻訳文を出力する。

以上の従来技術においては、特に手続4において重大な欠点がある。それは、木を木状パターンで覆う際に、採用可能な木状パターン間の対立可能性について何らの考慮も払われていないため、

すように、パターンE、F、Iによる3通りの覆い方が可能である。これをパターンの対立可能性という。従来技術においては、この対立可能性を考慮していなかったため、図20～図22のいずれを選択するか定かでない。このため、前述のように、下位に手続が進んだ際、覆うことのできるパターンが無かったり、不適切な目的言語を与えするという現象がしばしば発生していた。

(発明の課題)

本発明は、上述した従来技術の欠点を解決するためなされたもので、パターンの対立可能性を考慮し、与えられた木を最適に覆うことのできる言語木パターン被覆方式を提供することを目的とする。

(発明の構成および作用)

以下、先ず本発明の原理につき説明する。説明に当っては、与えられた木を適用可能なパターンで重複することなく、かつ間隙なく被覆する手続を全被覆手続、この全被覆手続を構成する各段階での手続を単位手続、この単位手続を構成するパ

ターンによる被覆を被覆手続と呼ぶ。また被覆規則は従来技術と同様、「各パターンは唯一の最上位節点を持ち、この最上位節点は上位パターンの最下位節点の一つを構成する」ものとする。このとき、あるパターンの最上位節点が、上位パターンに含まれない状態を間隙があるといい、上位パターンの最下位節点より上位節点の一つを構成する状態を重複するという。

なお、対立可能性は、上述したパターンの対立可能性の他に、従来技術の手続3、すなわち、バージングの段階でも下位節点を上位節点に結ぶ際に、その結び方に対立可能性が発生することが考えられる。この場合にはバージング段階で複数の木が発生し、本発明の被覆手続に対して複数の木が与えられる。

本発明は、複数の木が与えられても全被覆が完成できるものであるが、原理的なものとして1つの木が与えられた場合の被覆方式について先ず説明する。上述と同様、図7の与えられた木を図8～19のパターンを用いて覆うこととする。

目の単位手続を適用する。図20は、図24のごとく、パターンR、S、T、NおよびZを用いて被覆され、それぞれのパターンは対立可能性を有しない。これにより、図23の被覆木は図25の被覆木へと成長する。図25の節点[E]の下位には対立可能性が無いので対立可能性を示す小さな丸印は無い。

図21の被覆状態に第2回目の単位手続を適用すると、図26および27に示す2つの被覆状態が得られる。これは、パターンLおよびMの間に対立可能性が発生したからである。これにより、図25の被覆木は図28の被覆木へと成長するが、節点[L]と[M]の上には対立可能性を示す小さな丸印が発生している。

図22の被覆状態に第2回目の単位手続を適用すると、図29の対立可能性のない被覆状態を得、これによって被覆木は図30のものに成長する。

次に、図24、26、27および29の被覆状態に対し、第3回目の単位手続を適用する。この場合

最初の単位手続の適用によって、図20～図22に示すように3つの対立可能性が発生することは先に述べた。ここで、対立可能性の発生を示すために、被覆木なる表現方を導入することとする。すなわち、図20～図22を1つにまとめて図23のように表記することとする。図23において、[V]なる節点は被覆木の最上位節点であって木の形を整えるため導入されるもので特に意味はない。節点[V]の下位に接続された[E]、[F]および[I]の各節点は図20～図22に示されるごとく、木がパターンE、FおよびIで覆われることを示している。そして、節点[E]、[F]および[I]の上の小さな丸印は、それぞれの節点間に対立可能性があることを示している。すなわち、被覆木においては、小さな丸印の下位で同位のもは互いに入れ換えられるが、小さな丸印のない節点は入れ換えることができないことになる。

次に、各節点[E]、[F]および[I]に対応する被覆状態、すなわち、図20～22に第2回

対立可能性は発生せず、それぞれ図31、32、33および34の被覆状態が得られ、これに伴い図30の被覆木は図35の被覆木へと成長する。

次に、さらに、単位手続を適用する余地のある図32および34の被覆状態に対して第4回目の単位手続を適用する。この場合、図32に対しては図17および18のパターンXおよびYを用いて被覆が可能であり、図36の被覆状態が得られる。しかし図34については、パターンRおよびSを用いても上位パターンIとの間に間隔が生じ被覆規則に反することとなる。このことは、図34のようなパターンの被覆法、ひるがえって、図29および22の被覆法を選択すると、図7の与えられた木を図8～19のパターンを用いては完全に被覆できないことを意味する。このような場合、被覆木の被覆不可能な位置に⊗印を付して、これより下位の部分木の被覆は不可能であることを示す。

以上、単位手続を繰返し適用した結果、得られる被覆木は図37のものとなり、これを完成被覆

木と呼す。図 37 中、実線または鎖線で示した領域は、完成被覆木に占める図 25, 27, 28 および 30 の最終被覆状態の領域を示していて、領域中の数字が図番を指している。

これらの領域のうち、いずれかの領域が正しい目的言語を与えるための木であり、この正しい木を選び出す手続をトーナメント手続と呼ぶ。

以下、トーナメント手続について述べる。

トーナメント手続の適用を行うには、先ず図 37 の完成被覆について修正を加える必要がある。この修正は各パターンにパターン値を付加することであり、修正結果は図 38 となり、これをパターン値付完成被覆木と呼す。(図では、説明の都合上、各節点に番号を付してある) 図 38 中、大括弧 [] 中のパターン記号の横の数値がパターン値である。このパターン値は、翻訳上の経験則によって決められるが、固定値であっても、翻訳文のジャンルによって変化させてもよい。これについては後に詳説する。

なお、パターン値は、図 8 ~ 18 のパターンに

対して予め定まるものであるから、被覆木を作成する各単位手続の中でパターン値を付加しておけば上記修正は必要ないこととなる。

上述のように、トーナメント手続は図 38 のパターン値付完成被覆木において行われる。このトーナメント手続は、各節点にコスト関数値を導入し、コスト関数値が最大のものを選ぶという操作である。このコスト関数値としては、例えば図 7 の与えられた木が自然言語のパーズ木であれば、ある節点のコスト関数値は、その節点を含めて、その節点の下位に接続される全ての節点のパターン値の総和に選ぶのが妥当と思われる。これをここでは下位総和コスト体系と呼ぶこととする。後に、考えられる他のコスト体系について述べるが、先ず、下位総和コスト体系におけるトーナメント手続について説明する。なお、このコスト体系においては、被覆が不可能であることを示す⊗印の節点と、この節点から上に向かって対立可能性を示す小さな丸印に至るまでの全ての節点のコスト関数値は 0 であるとする。

図 38 中、まず節点 14 および 19 についてそのコスト関数値を計算すると、それぞれ 13 および 14 となり、より高いコスト関数値を有する節点 19 がここでのトーナメント手続の勝者に選ばれる。次に節点 2, 10 および 23 のコスト関数値を計算してみる。節点 2 のコスト関数値は、28 となる。節点 10 に関しては、下位に対立可能性ある節点 14 と 19 を有するが、このような場合は、上述した勝者系列のパターン値を上位節点のコスト関数値計算に用いることとすれば、節点 10 のコスト関数値は 27 となる。一方、節点 23 のコスト関数値は上述したところにより 0 である。

したがって、節点 2, 10, 23 のうちからは、節点 2 がトーナメント手続の勝者として選ばれる。勝者の系列を本発明の言語木パターン被覆方式の出力とするので、節点 2 が勝者であることは、図 31 に示す覆い方が出力されていることになる。

ここで他の可能なコスト体系について述べる。その 1 つは、ある節点のコスト関数値がその節点

より上位の節点のパターン値に依存するようなコスト体系であり、また、各パターンが複数個のパターン値をもち、これらの関数として関数値を与えるというコスト体系等である。

特に、複数個のパターン値を持つコスト体系の例としては、1 つのパターンが 2 つのパターン値を持つコスト体系に於て第 1 のパターン値は通常の下位総和コスト体系のパターン値に選び、第 2 のパターン値として値の大きいほど優先度の高い離散的な値を用い、かつ、対立可能性のうちの 1 つの選択に於ては第 1 のパターン値の系列の下位総和形のコスト関数値が最大のものではなく対立可能性の○印の直下位の節点に対応するパターンが持つ第 2 のパターン値が最大の節点を選ぶことを行ない、同一の第 2 のパターン値を持つ節点が 2 つ以上ある場合にのみ第 1 のパターン値に関連する下位総和形のコスト関数値が最大の節点を選ぶ方法も考えられる。

さらに各パターンが 3 つのパターン値をもつコスト体系も考えられ、上記 2 つのパターン値のほ

かに、優先度に関するパターン値を与え、このパターン値がある程度大きい場合には、このパターンのスパン（そのパターンが有する下位の木の最下位節点）と少しでも重なる他のパターンを全て排除する方法などである。

以上述べたある1つの与えられた木に対するあるコスト体系、およびある与えられたパターン群のもとでのコスト関数値を最大にする意味での最適な覆い方の選択法は自然言語処理の際の最適な木の変換単位を選択に有効に用いられる。すなわち原言語（翻訳を受ける元の言語）を表現する木に対しパターンによる被覆を行ないこの被覆パターン毎に目的言語（翻訳先の言語）の木に変換する事を行なえば容易に言語間の木の変換が行われる。パターン毎の変換に関しては「特願昭57-83961 機器翻訳方式」に詳細に記されているがこれにはパターンによる最適被覆法についてふれられていない。本発明はこの最適被覆法を選定する手段を提供するものであり、従来変換単位を与えられた木の情報より適切な方針なく選ん

でいたために誤変換が発生しがちであったのに対し本願の方法を併用しその際前述した下位総とコスト体系を用いると誤変換を減らすことが出来ることが判明している。

以上が、本発明を構成するための第1の方法である。

次に、複数の与えられた木を、ある与えられたコスト体系と与えられたパターン群のもとより、適切に覆われる木を選ぶ方法について説明する。説明例としては、これまで述べてきたものを含む。

今、図7に示す与えられた木とともに、図39に示す木が与えられ、さらに図8～19のパターン群に図40のパターンE'と図41のパターンF'が加えられたとする。なお、パターンE'のパターン値は18、パターンF'のパターン値は15であるとする。

図39の与えられた木に対するパターン値付完成被覆木は、前述の第1の方法の手法により図42に示すものになる。また、図38のパターン値付

完成被覆木は、図8～19のパターンに図40～41のパターンが加わったパターン群を用いて被覆手続を行った時、単に図40～41のパターンが適用されなかったものと考えてよい。したがって、図39および図38の双方のパターン値付完成被覆木は全く同一条件のもとで得られたものとみなせる。

これら、2つのパターン値付完成被覆木に対して、ここで完成被覆木融合法と呼ぶ手法を適用する。この手法は、「完成被覆木の上位節点より順にパターンマッチングを行い、マッチングの取れた部分は単一の部分にまとめ、マッチングの取れなかったものは対立可能性を示す小さな丸印の下位に置く。」という規則により1つの木を構成するものである。このようにして構成した木を「融合完成被覆木」と称し、本例では図43に示すものを得る。図43では、図38と42の節点1、および23より下位の部分はマッチングがとれたので単一部分にまとめられ、図38と図42それぞれの節点2および節10の下部部分はマッ

グが取れなかったことを示している。

このようにして得られた図43の融合完成被覆木に対して前述と同様トーナメント手続を施す。この結果、図中線で囲まれた系列が、最も大きいコスト関数値32を示し、この系列が選択され、出力される。この出力に対応する被覆状態は図44で、これは、図39に示す与えられた木に対応し、図7に示す与えられた木により最適に覆われることができることを示している。

このように融合完成被覆木は与えられた木のそれぞれに対する完成被覆木は前述した完成被覆木融合法で融合して得ることが出来るが、これに対し与えられた木の表現に於て前述の対立可能性の表現法を導入し複数の与えられた木を1つの木で表現することを行ない、これにより融合完成被覆木を直接得る方法が考えられる。

例えば図7及び図39の2つの木は対立可能性を導入することにより図45に示す1つの木にまとめられる。このような対立可能性を含む与えられた木に対し前述の単位手続を含む全被覆手続を

適用すると融合完成被覆木を直接構成できる。すなわち図45の木より図43の木を直接構成できる。この場合は単位手続は与えられた木に於ける対立可能性の組合わせの全てに適用しなければならないことは言うまでもない。この方法による融合完成被覆木構成法を前述した完成被覆木融合法に対比させて「直接法」と称することとする。

以上述べた融合完成被覆木の2つの構成法から明らかなように融合完成被覆木は複数の与えられた木に対する同一パターン群による覆い方の全てを示したものである。従ってこの融合完成被覆木に於てあるコスト体系にてトーナメント手続を行って得た最良の被覆法は与えられた複数の木の中で最も適切に覆われることの出来る木に於ける覆い方を示したものである。すなわち融合完成被覆木から求められた最良の被覆法は最良の被覆法を選定すると同時に最も適切に覆われる事の出来る木をも選定しているのである。この最も適切に覆われる事の出来る木は自然言語解析に際し、言語特に前述の原言語の木を表現するに最適な木であ

ることが判明しているのはこれは前述の「特願昭57-83981 機械翻訳方式」にて用いる原言語の木として最適に用いることが出来る。

前述した与えられた1つの木を最適に覆う方法が本発明を構成する第1の方法であるのに対し、ここで述べた最適に覆われる事の出来る木をその覆われ方と共に求める方法は本発明を構成する第2の方法である。

以上、本発明の原理について説明した。以下、本発明の一実施例について説明する。本実施例は上述した本発明を構成する第1の方法に対応するものである。

まず、木の直列記述法について述べる。この記述法は、図7に示されるような2次元的に表現された木を、電気信号として扱う場合の1次元的な表現方法である。

木は基本的に次式のように表現される。

$$\text{木} = \text{親節点}(\text{子節点1} \sqcup \text{子節点2} \sqcup \text{子節点3} \sqcup \dots) \quad (1)$$

式1は図46のような節点の関係を示すもので

ある。すなわち、カッコの左外側に親節点（上位節点）をおき、カッコ内に子節点（直下節点）をスペース \sqcup で区切り、木における位置関係を保ちながら並べる。この直記法によれば、図7に示される木は次式のように表わされる。

$$A(B(G(R) \sqcup H(S)) \sqcup C(T) \sqcup D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))) \quad \text{--- (2)}$$

また、パターンは次式で表現される。

$$\text{パターン} = \langle \text{パターン値} \rangle \text{親節点}$$

$$(\text{*子節点1} \sqcup \text{*子節点2} \sqcup \dots) \quad (3)$$

ここで、アスタリスク記号*は、この記号が付された節点の下位構造が任意であることを示す。したがって、アスタリスク記号*の付されない子節点は終端節点とならなければならない。

式(3)によれば図8のパターンEは次式で示され、その関係は図47である。図47の形式の記述を式表現と呼ぶ。例えば図47はパターンEの式表現である。

$$\langle 17 \rangle A(B(*G \sqcup *H) \sqcup *C \sqcup D(*J \sqcup *K)) \quad \text{--- (4)}$$

また、図14のパターンRは次式で示される。

$$\langle 1 \rangle G(R) \quad \text{--- (5)}$$

この記述法による図8～図19までのパターンを表1に整理しておく。

パタン	表 現	関係図
E	$\langle 17 \rangle A(B(*G \sqcup *H) \sqcup *C \sqcup D(*J \sqcup *K))$	図 47
F	$\langle 10 \rangle A(B(*G \sqcup *H) \sqcup *C \sqcup *D)$	48
I	$\langle 15 \rangle A(*B \sqcup *C \sqcup D(*J \sqcup *K))$	49
L	$\langle 5 \rangle D(*J \sqcup *K)$	50
M	$\langle 11 \rangle D(J(*P \sqcup *Q) \sqcup *K)$	51
N	$\langle 5 \rangle J(*P \sqcup *Q)$	52
R	$\langle 1 \rangle G(R)$	53
S	$\langle 1 \rangle H(S)$	54
T	$\langle 1 \rangle C(T)$	55
X	$\langle 1 \rangle P(X)$	56
Y	$\langle 1 \rangle Q(Y)$	57
Z	$\langle 1 \rangle K(Z)$	58

表 1

次に図59図に構成例を示し、これについて説明する。

図において、1は木をパターンにより被覆するパターン適用部、2は単位手続を行いパターン値付完成被覆木を構成する結合部、3は下位総和コスト体系にもとづいてトーナメント手続を行うトーナメント部、4はクロック発生器である。

先ず、パターン適用部1を図80に示し、説明する。パターン適用部1は、表1に示したパターンを収容するパターン辞書メモリ11と、与えられた木を記憶する部分トリメモリ12と、与えられた木とパターンとを照合することによって被覆手続を行うパターン適用演算部13と、パターン適用演算で発生し、次の単位手続の対象なる部分木を蓄積する部分トリメモリ14と、演算結果として、ある木を1つのパターンで完全に覆うことが出来た場合、その結果を記憶するターミナルトリメモリ18と、1つのパターンで完全に覆うことが出来なかった場合、その被覆状態を記憶するフラットトリメモリ15からなる。

パターン辞書メモリ11は、例えば M_1 段のシフトレジスタを N 個からなるような M_1 行 N 列のメモリで M_1 はパターンの数、 N は木の最大長によって決められる。部分トリメモリ12、蓄積トリメモリ14、フラットトリメモリ15およびターミナルトリメモリ18は基本的に同じ構成で、 M_2 行 N 列のメモリである。 M_2 は木の

取り扱い量によって決められる。パターン辞書11をはじめ上述のメモリは全て巡回読み出しと非巡回読み出しの両方が可能なメモリである。

部分トリメモリ12は、外部より図7に相当する与えられた木が収容される。この与えられた木はもちろん上述した直列記述法による電気信号であつてゐる。いま、部分トリメモリ12のある行に図7で示された与えられた木、すなわち、式(2)で表わされる木が収容されていて、この与えられた木の処理の順番になったとする。

先ず、部分トリメモリ12から与えられた木が、パターン辞書メモリ11から表1に示すパターンEがパターン適用演算部13に取り込まれる。適用演算部13の演算手法を図61に示している。すなわち、 $\langle \rangle$ 括弧はそのまま出力することとして、1文字毎に一致不一致を検出する。そして、アスタリスク*の付いた節点との一致が検出されると、このアスタリスク*の付いた節点は下位系列が任意であるので、与えられた木の節点にシャープ記号#を付して、その下位系列をそ

のままパターンに代入する。例えば、図61で、与えられた木の節点GとパターンEの節点Gとの一致が検出された場合、パターンEの節点Gには、その前にアスタリスク*があるので、*Gの後に括弧を追加し、この括弧内に、与えられた木の節点Gにシャープ記号#を付するとともにその下位系列を代入する。図から判るように、パターンEの節点*Gの後に(*G(R))が追加されている。同様な演算が、節点H、C、JおよびKについて行われている。

この演算により、パターンEに追加された(*G(R))は、パターンEによって覆うことのできなかった部分木であり、これは、他のパターンによって覆われるべきものであるから、演算の際にG(R)の形で蓄積部分トリメモリ14に記憶される。図61の例で、蓄積部分トリメモリ14に記憶されるのは、G(R)、H(S)、C(T)、J(P(X))、Q(Y))およびK(Z))である。

図61の演算で得られた結果は、フラットトリメモリに移されるが、これについては後述

する。

パターンEに関する演算が終了すると、パターン辞書メモリ11から、表1に示す次のパターン、すなわちパターンFがパターン適用演算部13に読込まれ、与えられた木とパターンFとの間で前述の手法で演算が行なわれる。この場合の演算結果は次式のものとなる。

$$\langle 10 \rangle A(B(*G(*G(R)) _ H(*H(S))) _ C(*C(T)) _ D(*D(J(P(X)) _ Q(Y))) _ K(Z))) \text{ --- (6)}$$

このとき、前述と同様、パターンFで覆われなかった部分木は、蓄積部分トリメモリ14に移されるが、すでに記憶されている部分木は除かれる。したがって、パターンFに関して発生する部分木はD(J(P(X))、Q(Y))、K(Z))である。

次に、パターン辞書メモリ11より、パターンIがパターン適用演算部13に読込まれ、与えられた木とパターンIとの間の演算が行われる。この演算結果は次式に示すものである。

$$\langle 15 \rangle A(*B(*B(G(R)) _ H(S))) _ C(*C(T)) _ D(*J(*J(P(X)) _ Q(Y))) _ K(*K(Z))) \text{ --- (7)}$$

またこの際に蓄積部分トリメモリ 14 に新たに記憶される部分木は $B(G(R) \sqcup H(S))$ である。

ここで、パターン適用演算部 13 の出力形式について述べておく。パターン適用演算部 13 の演算結果として、与えられた木の全体を 1 つのパターンで覆うことができた場合、その演算結果は確定されるのでターミナルトリメモリ 18 へ移し、1 つのパターンで覆うことのできない部分木を含む場合は、その演算結果は未確定なのでフラットトリメモリ 15 へ移すこととする。そして、図 81 に示した演算結果及び式 (), () の演算結果は全て、式 (2) で示される木についてのものであり、このことを表示するため、次式の形式をとる。

木 (演算結果 1, 演算結果 2 ----- ,
演算結果 n) --- (8)

したがって、以上の演算結果は式 2 を α 、81 の結果を β 、式 6 を γ 、式 7 を δ とすると次式の形式でフラットトリメモリ 15 に記憶される。

$\alpha(\beta\gamma\delta)$ --- (9)

$C(T)(\langle 1 \rangle C(T))$ --- (12)

$J(P(X) \sqcup Q(Y))(\langle 5 \rangle J(*P(*P(X)) \sqcup *Q(*Q(Y))))$
--- (13)

$K(Z)(\langle 1 \rangle K(Z))$ --- (14)

$D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))(\langle 5 \rangle D(*J(*J(P(X) \sqcup Q(Y))) *K(*K(Z))) \langle 11 \rangle D(*P(*P(X)) \sqcup *Q(*Q(Y)) \sqcup *K(*K(Z))))$
--- (15)

ところで、残る $B(G(R) \sqcup H(S))$ の部分木についてはいずれのパターンを用いても覆うことができない。このような場合は、0 値をもつパターンが適用されたとして次式のように表現される。

$B(G(R) \sqcup H(S))(\langle 0 \rangle)$ --- (6)

式 (10) ~ 式 (15) の結果において、パターンによって覆われない部分木をもつものは式 (13) および式 (15) であり、これらの結果はフラットトリメモリ 15 に記憶され、他の確定された結果はターミナルトリメモリ 18 に記憶されていることとなる。

次に結合部 2 とトーナメント部 3 について説明する。

この式 (8) を図式化して図 82 に示しておく。

以降、表 1 に示すパターン I からパターン Z までは順次パターン適用演算部 13 に読み出され、与えられた木との間で演算が行われるが、これらの間には一致性がなく上述のような演算結果は生じない。

以上のように、パターン E からパターン Z までを対象とした演算が終了すると、次には、蓄積部分トリメモリ 14 に蓄積されている部分木について演算を行なうこととなる。ここまでに、蓄積された部分木について整理すると、 $G(R)$, $H(S)$, $C(T)$, $J(P(X) \sqcup Q(Y))$, $K(Z)$, $D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))$ 及び $B(G(R) \sqcup H(S))$ の 7 つの木である。これらの木が順次、蓄積部分トリメモリ 14 よりパターン適用部 13 に読み出され、パターン E ~ Z を対象に演算が行なわれる。その演算結果は、上述の出力形式と同じであって、それぞれ次式に示されるものとなる。

$G(R)(\langle 1 \rangle G(R))$ --- (10)

$H(S)(\langle 1 \rangle H(S))$ --- (11)

図 83 は、結合部 2 とトーナメント部 3 を示し、図において、21 は中間トリメモリ、22 は確定トリメモリ、23 は結合演算部、31 はトーナメント部である。中間トリメモリ 21、確定トリメモリ 22 はそれぞれパターン適用部 1 のフラットトリメモリ 15 とターミナルメモリ 18 に対応し、これらと全く同じ M_2 行 N 列の記憶回路である。

先ず、パターン適用部 1 のフラットトリメモリ 15 の記憶内容が中間トリメモリ 22 へ、ターミナルトリメモリ 18 の記憶内容が確定トリメモリ 22 へ移行導入される。中間トリメモリ 21 の内容は 1 行ずつ結合演算部 23 へ読み出されて、図 84 に示す演算が行なわれる。すなわち、いま、前述した $\alpha(\beta\gamma\delta)$ が演算部 23 に読み込まれたとすると、その先頭からシャープ記号 # の付いた部分木、すなわちまだパターンで覆われていない部分木を検出し、この木を確定トリメモリ 22 に記憶されている確定木によって被覆を行い、それによって同時にパターン値付け

を行なうものである。図 84 は $\alpha(\beta\gamma\delta)$ のうち β 部分が演算の対象となっている演算の途中図である。すなわち、演算規則は、例えば、 $*G(R)$ を見出しした場合、シャープ記号 # を除く $G(R)$ を対象に、これと一致性のある確定木 $G(R)(\langle 1 \rangle G(R))$ を確定トリメモリ 22 から探し出し、確定木 $G(R)(\langle 1 \rangle G(R))$ の最初の $G(R)$ を除く部分 $(\langle 1 \rangle G(R))$ を前記の $*G(R)$ に代入するものである。この演算により、 $\alpha(\beta\gamma\delta)$ は $\alpha(\beta'\gamma'\delta')$ と変形され、 β' 、 γ' 、 δ' はそれぞれ次式 (17)、(18)、(19) で示される形となる。

$$\langle 17 \rangle A(B(*G((\langle 1 \rangle G(R))) \sqcup *H((\langle 1 \rangle H(S)))) \sqcup *C((\langle 1 \rangle G(T))) \sqcup D(*J(*J(P(X) \sqcup Q(Y))) \sqcup *K((\langle 1 \rangle K(Z)))) \quad \text{--- (17)}$$

$$\langle 18 \rangle A(B(*G((\langle 1 \rangle G(R))) \sqcup *H((\langle 1 \rangle H(S)))) \sqcup *C((\langle 1 \rangle G(T))) \sqcup *D(*D(J(P(X) \sqcup Q(Y)) \sqcup K(Z)))) \quad \text{--- (18)}$$

$$\langle 19 \rangle A(*B((\langle 0 \rangle)) \sqcup *C((\langle 1 \rangle G(T))) \sqcup D(*J(*J(P(X) \sqcup Q(Y))) \sqcup *K((\langle 1 \rangle K(Z)))) \quad \text{--- (19)}$$

この演算結果において、その信号中にシャープ記号 # を含む場合には、まだ確定されていないので、中間結果トリメモリ 22 に再度格納され

る。

次に、中間結果トリメモリ 21 より、式 (13) で示される信号が結合演算部 23 へ導入され、演算を受ける。 $J(P(X) \sqcup Q(Y))(\langle 5 \rangle J(*P(*P(X)) \sqcup *Q(*Q(Y))))$ のうち $*P(X)$ 、 $*Q(Y)$ の部分が処理を受け、 $J(P(X) \sqcup Q(Y))(\langle 5 \rangle J(*P((\langle 1 \rangle P(X))) \sqcup *Q((\langle 1 \rangle Q(Y))))$ と変形される。この結果中には、シャープ記号 # がないのでトーナメント部 3 へ出力される。

トーナメント部 3 は、入力信号中に対立可能性、すなわち、2 つの部分木を含んでいる場合には、それぞれの部分木の下位総和コスト関数を求めて勝者を決定し、入力された木を確定する。この確定された木は確定トリメモリ 22 へ出力され、以降の結合演算に用いられる。上記 $J(P(X) \sqcup Q(Y))$ の場合には、対立可能性は発生していないので、トーナメント部 3 はそのまま確定トリメモリ 22 へ出力する。

次に式 (15) で示される $D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))$ については、対立可能性がある。式 (15) は

結合演算により次式に変形される。

$$D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))(\langle 5 \rangle D(*J((\langle 5 \rangle J(*P((\langle 1 \rangle P(X))) \sqcup Q((\langle 1 \rangle Q(Y)))) \sqcup *K((\langle 1 \rangle K(Z)))) \langle 11 \rangle D(J(*P((\langle 1 \rangle P(X))) \sqcup *Q((\langle 1 \rangle Q(Y)))) \sqcup *K((\langle 1 \rangle K(Z)))) \quad \text{--- (20)}$$

この結果はトーナメント部 3 で $\langle 11 \rangle$ で始まる文字列が勝者と判定され、次式の確定木として確定トリメモリ 22 へ格納される。

$$D(J(P(X) \sqcup K(Z))(\langle 11 \rangle D(J(*P((\langle 1 \rangle P(X))) \sqcup *Q((\langle 1 \rangle Q(Y))) \sqcup *K((\langle 1 \rangle K(Z)))) \quad \text{--- (21)}$$

以上の演算により、最初に中間結果トリメモリ 21 に記憶されていた木についての処理は全て終了し、次に中間結果トリメモリ 21 に収容されている $\alpha(\beta'\gamma'\delta')$ について再度結合演算が施される。この時点で式 (17) 及び (18) 中の $*J(P(X) \sqcup Q(Y))$ の部分と式 (18) 中の $*D(J(P(X) \sqcup Q(Y)))$ の部分に代入すべき確定木は以前の結合演算により確定トリメモリ 21 に準備されているので、それらの部分が代入変換される。そして、シャープ記号 # を含まない文字列となり、トーナメント部 3 に出力される。

トーナメント部 3 では、前述のように確部分木のコスト関数計算により勝者を決定し、 β' を選択することとなる。この結果はトーナメント部 3 より確定トリメモリ 22 に次式の形で出力される。

$$A(B(G(R) \sqcup H(S)) \sqcup C(T) \sqcup D(J(P(X) \sqcup Q(Y)) \sqcup K(Z))) \langle 17 \rangle A(B(*G((\langle 1 \rangle G(R))) \sqcup *H((\langle 1 \rangle H(S)))) \sqcup *C((\langle 1 \rangle G(T))) \sqcup D(*J((\langle 5 \rangle J(*P((\langle 1 \rangle P(X))) \sqcup *Q((\langle 1 \rangle Q(Y)))) \sqcup *K((\langle 1 \rangle K(Z)))) \quad \text{--- (22)}$$

この式 (22) が最終的な全被覆手続の結果であり、外部に出力されるものである。式 (22) を図式化したものが図 85 である。

なお、トーナメント部 3 の詳しい説明は省略するが、入力となる文字列中の右括弧及び左括弧の数を制御子として、各部分木に分離し、各部分木中の $\langle \rangle$ 括弧中のパターン値の総和を求めれば、最適な被覆状態を有する木を選択することができる。

以上詳細に本発明の一部構成する 1 つの与えられた木に対する最適な被覆方法に関する実施例に

ついて述べて来た。次に本発明を構成する第2の方法である与えられた複数の木に対し最適な覆われ方をする木をその覆い方と共に得る方法の実現法について述べる。

今迄述べて来た実現例に於てはパターン適用部PAPにその出力として左括弧(の左側にある木又は部分木に対し、そのパターンによる覆い方の全てを括弧で囲んで列挙した形式の出力を発生する。図40に示したような対立可能性表現を含む木又は部分木を左括弧(の左側に置き括弧で囲まれた部分に上記木又は部分木の対立可能性の全ての組合せに対する全ての覆い方を列挙して示すように今迄説明したパターン適用部を変形することにより上記複数の与えられた木を取り扱う本発明を構成する第2の方法が実現できる。上記パターン適用部PAPの変形はこれまで述べたパターン適用部の前段に對立可能性表現を持つ木状構造を入力しその対立可能性の全ての組合せを出力として発生する部分を追加すればよくこの部分の構成法は自明であるのでこれ以上ここでは述べない。以

上述べたように複数の与えられた木に対し最適に覆われる木をその覆い方と共に発生する第2の方法が実現できるがここで述べた方法は本願冒頭で述べた本願を構成する第2の方法に於ける2つの主な融合完成被覆木構成法のうち直接法に対応するものである。

なお本発明の方法が縮退した形式として与えられる木の全部ではなくその一部を最適に覆う方法を発見する方法については、与えられた木の該当する部分を与えられた木としてこれに本発明のこれ迄述べて来た手法を適用すれば良いのでこれについてもこれ以上ふれない。

(発明の効果)

以上詳細に説明したように本発明は与えられた木と与えられたパターンを用いて与えられたコスト体系のもとで最適に覆う方法並びに与えられたいくつかの木につき上記手法を適用し最適に覆われることの出来る木をその覆い方と共に求める方法を提供するものであり、例えば「特願昭57-83981 機械的翻訳方式」のごとき与えられた木と

パターンが用いられ、しかも与えられた木が広々にして不適切な木を含む複数のものである場合に於ける自然言語解析に際して非常に有効に動作するものである。

4. 図面の簡単な説明

図1は従来技術に於ける入力文に対する節点の図、

図2は従来技術に於ける品詞確定に際する構造の図、

図3は従来技術に於ける言語木の図、

図4は従来技術に於ける木状パターンの図、

図5は従来技術に於ける木状パターンを用いた被覆状況の図、

図6は従来技術に於ける目的言語木の図、

図7はパターンで覆うべく与えられた木の図、

図8～図19はパターンの図、

図20～図22は被覆状態の対立可能性を示す図、

図23は最初の単位手続き適用後の被覆木の図、

図24は図20の構造に更に単位手続を適用し

て得た被覆状態の図、

図25は図20の被覆状態に第2回目の単位手続を適用した後発生する被覆木の図、

図26～図27は図21の構造に更に単位手続を適用して得た被覆状態の図、

図28は図21の被覆状態に第2回目の単位手続を適用した後発生する被覆木の図、

図29は図22の構造に更に単位手続を適用して得た被覆状態の図、

図30は図22の被覆状態に第2回目の単位手続を適用した後発生する被覆木の図、

図31～図34は第3回目の単位手続を適用して得た被覆状態の図、

図35は第3回目の単位手続を適用した後発生する被覆木の図、

図36は図32の被覆状態に第4回目の単位手続を適用して得た被覆状態の図、

図37は完成被覆木の図、

図38はパターン値付完成被覆木の図、

図39はパターンで覆うべく与えられた他の一

つの木の図、

図 40 ～ 図 41 は追加パターンの図、

図 42 は図 39 の木に対するパターン値付完成被覆木の図、

図 43 は融合完成被覆木の図、

図 44 は最適な被覆状態を示す図、

図 45 は対立可能性を用いて木をまとめる手法を示す図、

図 46 は式(1)に対応する木の図、

図 47 はパターンEの式表現の図、

図 48 はパターンFの式表現の図、

図 49 はパターンIの式表現の図、

図 50 はパターンLの式表現の図、

図 51 はパターンMの式表現の図、

図 52 はパターンNの式表現の図、

図 53 はパターンRの式表現の図、

図 54 はパターンSの式表現の図、

図 55 はパターンTの式表現の図、

図 56 はパターンXの式表現の図、

図 57 はパターンYの式表現の図、

図 58 はパターンZの式表現の図、

図 59 は本発明の構成例の図、

図 60 はパターン適用部の図、

図 61 は適用演算部の演算手法を示す図、

図 62 は図 61 の構造を図式化して示す図、

図 63 は、結合部及びトーナメント部を併せ示す図、

図 64 は結合演算部に於ける演算手法を示す図、

図 65 は電子的手段により得た全被覆手続結果の図である。

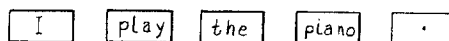
特 許 出 願 人

国 際 電 信 電 話 株 式 会 社

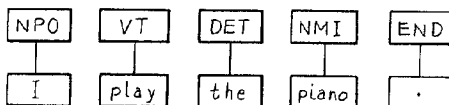
特 許 出 願 代 理 人

弁 理 士 山 本 恵 一

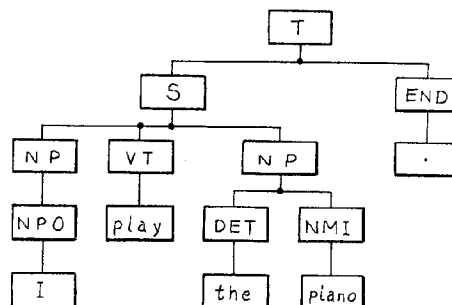
第1図



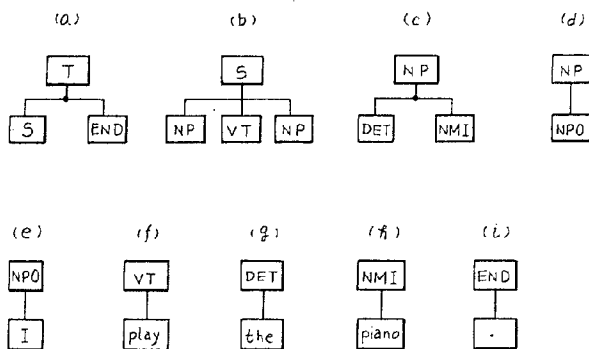
第2図



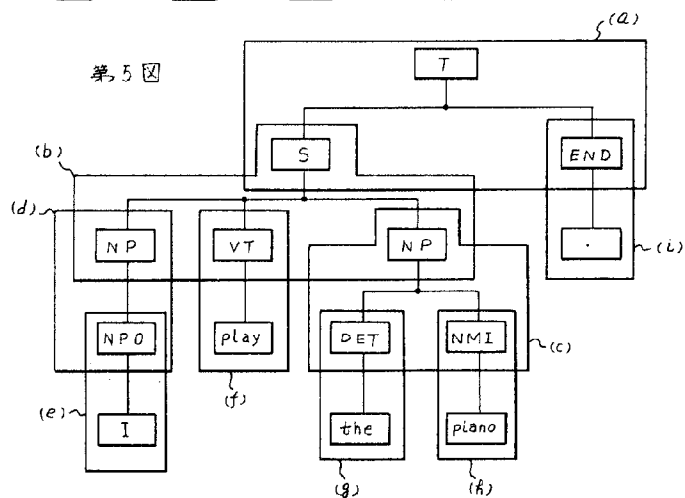
第3図



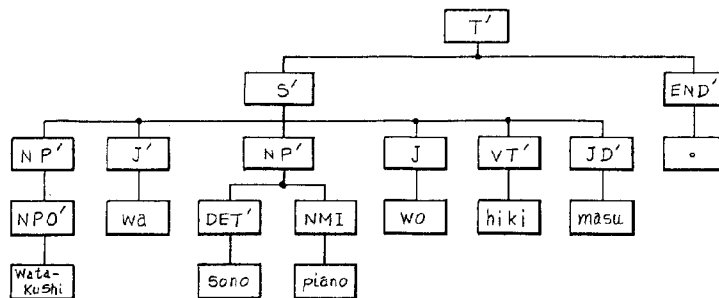
第4図



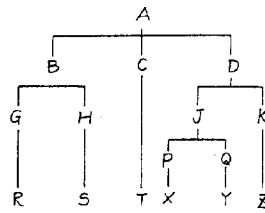
第5図



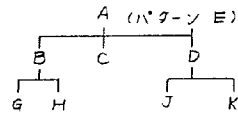
第6図



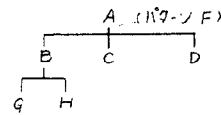
第7図



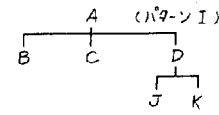
第8図



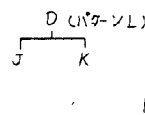
第9図



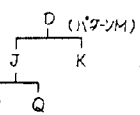
第10図



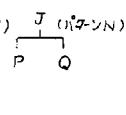
第11図



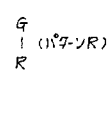
第12図



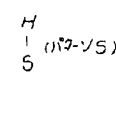
第13図



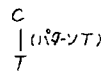
第14図



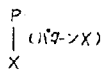
第15図



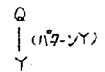
第16図



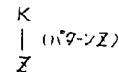
第17図



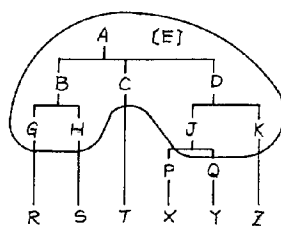
第18図



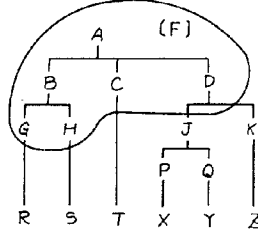
第19図



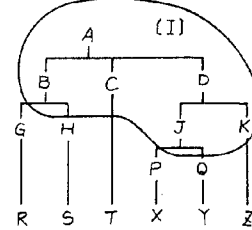
第20図



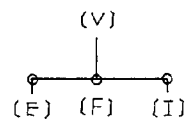
第21図



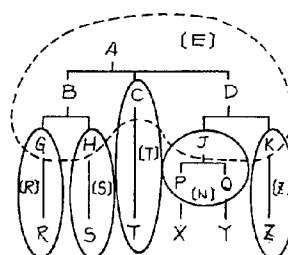
第22図



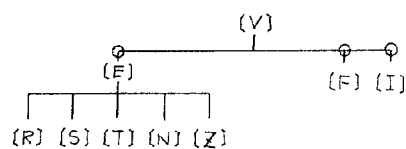
第23図



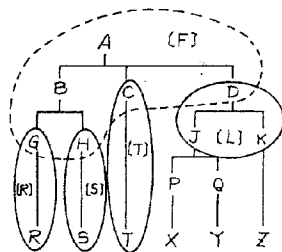
第24図



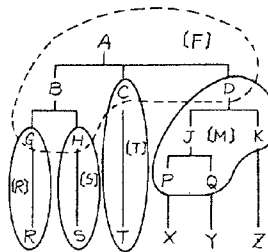
第25 図



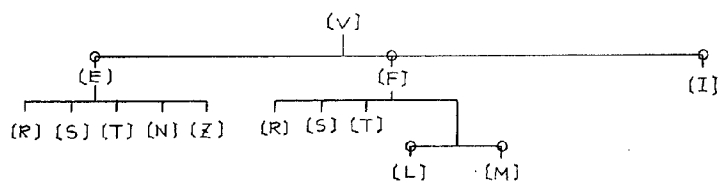
第26 図



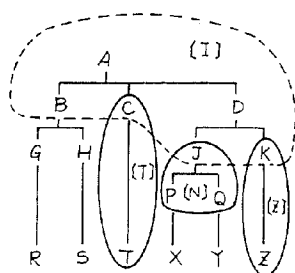
第27 図



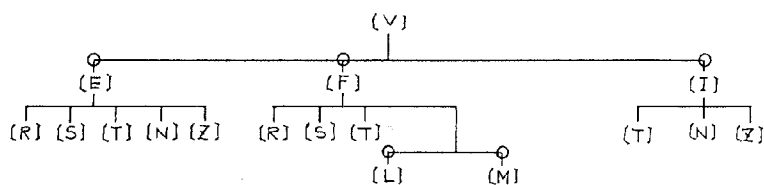
第28 図



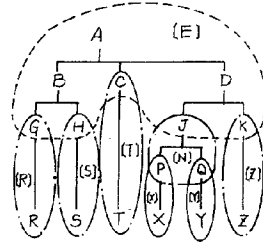
第29 図



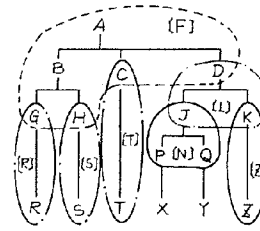
第30 図



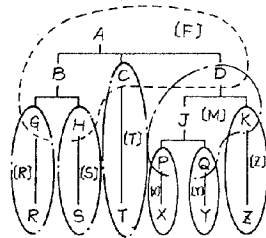
第31図



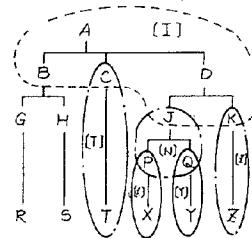
第32図



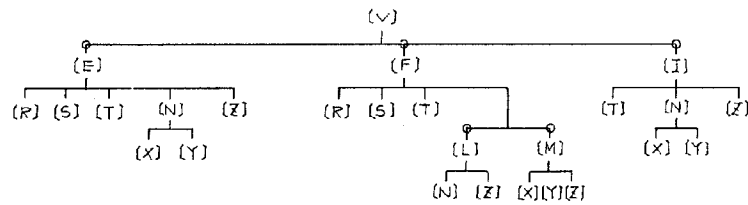
第33図



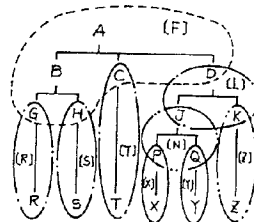
第34図



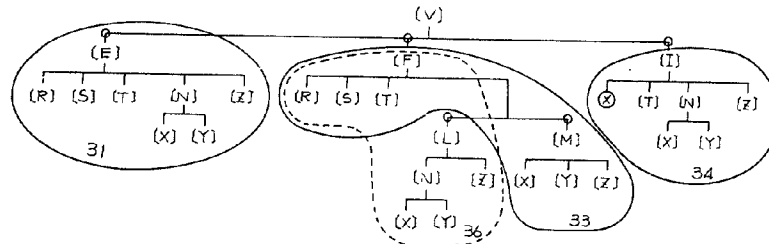
第35図



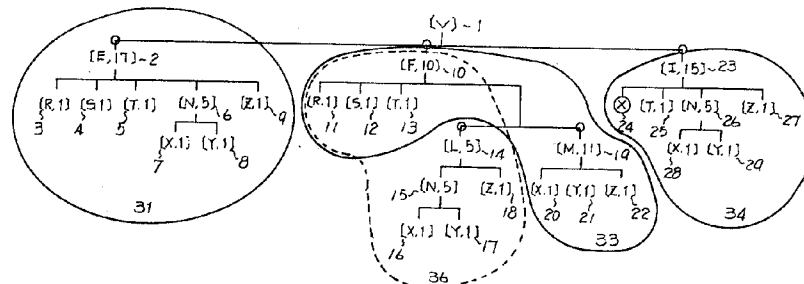
第36図



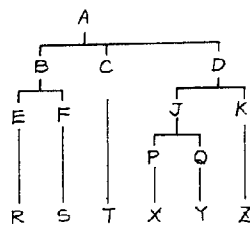
第37図



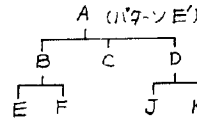
第38図



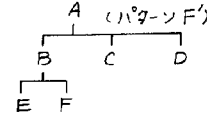
第 39 図



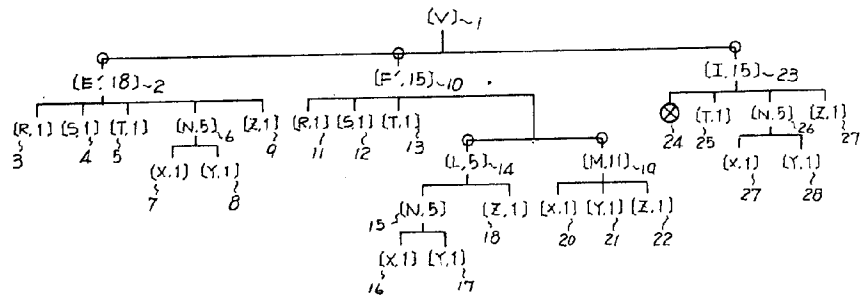
第 40 図



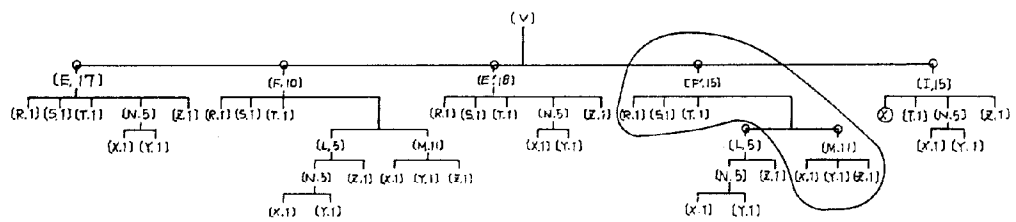
第 41 図



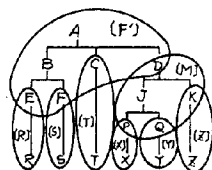
第 42 図



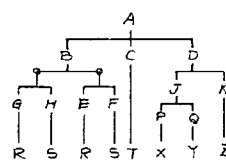
第 43 図



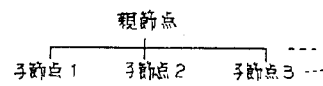
第 44 図



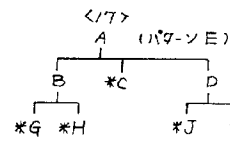
第 45 図



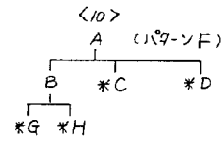
第46図



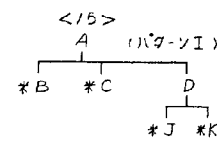
第47図



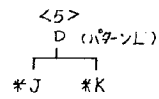
第48図



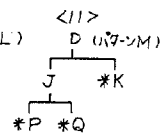
第49図



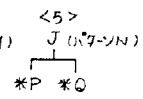
第50図



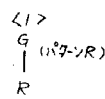
第51図



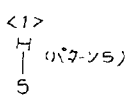
第52図



第53図



第54図



第55図



第56図



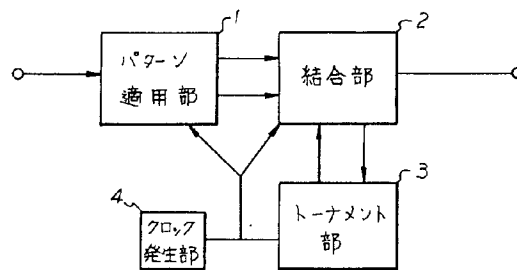
第57図



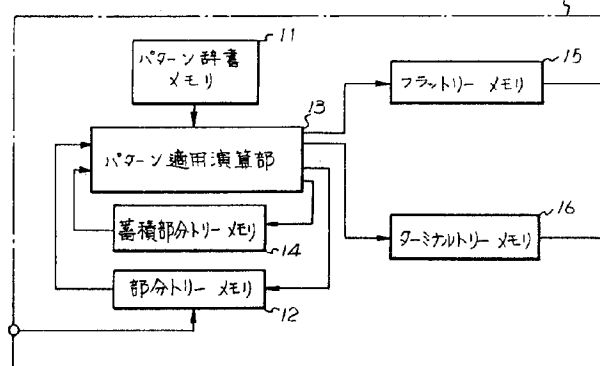
第58図



第59図



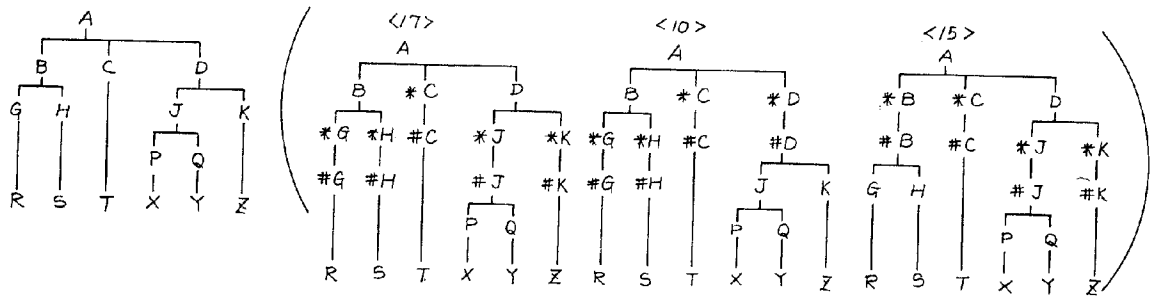
第60図



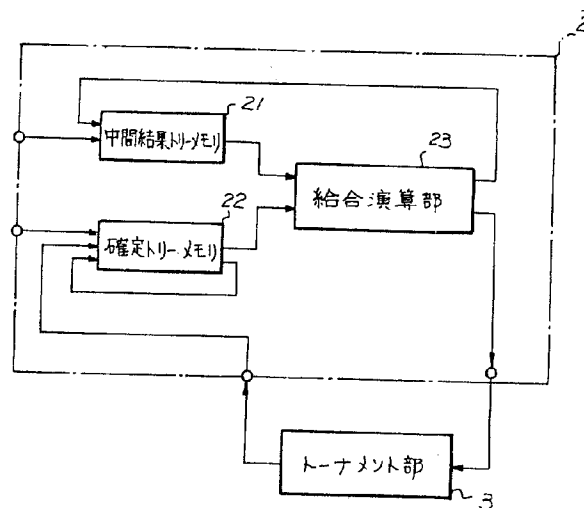
第 61 図

与えられた木: $A(B(G(R) \cup H(S)) \cup C(T) \cup D(J(P(X) \cup Q(Y)) \cup K(Z)))$
 パターン E: $\langle 17 \rangle A(B(*G \cup *H) \cup *C \cup D(*J \cup *K))$
 結果: $\langle 17 \rangle A(B(*G(\#G(R)) \cup *H(\#H(S))) \cup *C(\#C(T)) \cup D(*J(\#J(P(X) \cup Q(Y))) \cup *K(\#K(Z))))$

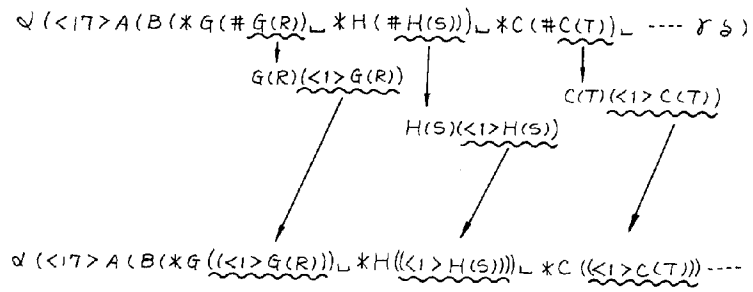
第 62 図



第 63 図



第64図



第65図

